Zeyu Li

3270 Sawtelle Blvd. Apt. 203, Los Angeles, CA, 90066

(+1) 424-386-8519 | lizeyu_cs@foxmail.com | zyli93 | zeyu-li-ucla

Education

University of California, Los Angeles (UCLA) *Los Angeles, CA, USA*

Ph.D. Student in Computer Science, supervised by Prof. Wei Wang, GPA: 3.77/4 *Sept. 2016 - PRESENT*

Research Interest: Network Embedding, Natural Language Processing, Data Mining, and Machine Learning

Harbin Institute of Technology (HIT) *Harbin, Heilongjiang, China*

B.S. in Computer Science, GPA: 87.86/100 *Aug. 2012 - Jun. 2016*

Publication

Repairing Data through Regular Expressions *VLDB 2016*

Zeyu Li, Hongzhi Wang, Wei Shao, Jianzhong Li, Hong Gao *Delhi, India*

Selected Projects

Personalized Domain Expert Recommendation on QA Websites *UCLA*

Graduate Student Researcher | *Numpy, PyTorch* *Aug. 2017 - Present*

- Learnt the similarity between users from user network structure information and accepted-answer ranking information to provide personalized and best answering candidates for recommendation given the question proposed by a specific user.
- Jointly optimized network embedding similarity loss between users and ranking loss amoung answers. Used Bidirectional LSTM for learning the representations of question contents.

All Case Reports with Open Biomedical Annotation Terms (ACROBAT) *UCLA*

Graduate Student Researcher | *Java* *Jul. 2017 - Present*

- To implement an automatic case report annotation tool to identify the typed medical entities, relationships, and temporal information to construct a graphical representation of a case report describing the treatment-disease development. Learn the representation of heterogeneous graphs for achieving high accuracy similarity search query and finding undocumented relationships between medical entities.
- Worked on Temporal Expression Extraction by introducing new hand-crafted rules to SUTime and collaborating with an Regex section.

Forecasting the best answers in community-based question answering services *UCLA*

Course Project | *Numpy, Sci-kit Learn* *Apr. 2017 - Jun. 2017*

- To automatically predict the best answer of questions in community-based question answering forum, such as Stack Overflow. Four types of features are employed: statistical and meta features, social network, user expertise features, and natural language features.
- Used LambdaMART, one of the state of the art learning to rank models, to learn the answers' ranking and select the highest as the predicted "best answer"
- Empirical results show that proposed model outperforms all baselines and enhances the 7% of the average P1.

Research of Basic Theory and Critical Technique on Big Data Usability *HIT*

Undergraduate Student Researcher *Mar. 2015 - Jun. 2016*

- Focused on developing data quality concepts and data cleaning algorithms, especially on regular-grammar-constrained data repairing techniques through pattern regular expressions.
- Proposed Regular-expression-based Structural Repair (RSR) algorithm which is a dynamic programming method with time complexity $O(nm^2)$ and space complexity $O(nm)$.
- Combined the RSR algorithm (for structural repair) with the value repair algorithm, a selection between association rule mining and edit distance. Constructed whole repairing methodology. Conducted excessive experiment to prove the effectiveness and efficiency.

Honors & Awards

2016	**Graduate Dean's Scholar Award**	*Los Angeles, U.S.A*
2016	**VLDB Travel Grant** VLDB 2016	*Dehli, India*
2016	**Excellent Graduate Student of Harbin Institute of Technology** (Top 10%)	*Harbin, China*
2016	**Outstanding Undergraduate Graduation Thesis** (Top 3%)	*Harbin, China*
2014	**National Scholarship** (Top 1%)	*Harbin, China*
2013	**CASIC (China Aerospace Science & Industry Corp.) Scholarship** (Top 3%)	*Harbin, China*